Kabayaki for Linux

辞書管理ツール取扱説明書

version 1.2.1

目次

はじめに **5**
ツールについて **5**
1章 **9**
文字と品詞 **9**
使用可能な文字 11
登録できる品詞 13
2章 **15**
辞書の編集 その1 **15**
単語の個別登録 17
単語追加登録 メニュー 17
単語の一括登録 22
単語一括登録メニュー 22
単語の削除 26
3章 **31**
辞書の操作 **31**
単語の検索･一覧表示 33
登録されているすべての単語を一覧する場合 33
登録した単語の頭文字で検索し一覧する場合 34
単語の読みで検索する場合 35
単語の更新 37

はじめに

ツールについて

たとえば, 法律に関する文書を Kabayaki にて扱う場合に, 裁判官, 訴訟, 税理士などの単語が登録されていなかったとします. (実際には登録されています)この状態で, Kabayaki にて「裁判官」で検索をかけた場合, 文中の「裁判官」という単語を正しく取得できない場合があり, 結果として望む検索の結果を取得できない可能性があります. また, 望む検索結果を取得できた場合でも, Kabayaki が「裁判官」をひとつの単語であると認識できていないので, 「裁判」や「裁判所」, 「警察官」(官の字でマッチしています)などの含まれる文書も検索結果に含まれてしまい, 結果として検索精度が落ちてしまう可能性があります. Kabayaki は一般的に用いられるほとんどの単語を網羅していますが, 特定の業界にのみ通じる語句や, Kabayaki の辞書が作成された時点以降に新たに生まれた単語に関してはサポートされていないものもあります. これらの理由による検索精度低下への対策として, Kabayaki ではユーザが自由に単語を登録し検索精度を高める機能が提供されています. それが, ここで解説する辞書管理ツールです. 辞書管理ツールを用いて専用の辞書登録を行えば, 医療業界, 出版業界, 法曹界, コンピュータ業界など, 専門性の高い分野における検索の精度を高めることが可能です.

1章
文字と品詞

辞書登録ツールで登録できる文字と品詞には制限があります．この章では，その使用可能な文字と品詞について解説します．

使用可能な文字

Kabayaki 辞書管理ツールにおける，使用可能な文字は，各フィールドに依って異なります．

新規登録時の「単語」フォーム

漢字，平仮名，片仮名，英数字(全角のみ)

新規登録時の「読み」フォーム

片仮名のみ

更新時の「単語」フォーム

漢字，平仮名，片仮名，英数字(全角のみ)

更新時の「読み」フォーム

片仮名のみ

一括登録用 CSV ファイルの，単語用カラム

漢字，平仮名，片仮名，英数字(全角のみ)

一括登録用 CSV ファイルの，読み用カラム

片仮名のみ

一括登録用 CSV ファイルの，品詞用カラム

『8.品詞の説明』にある品詞のみ

一文字でも誤りがあるとエラーが表示されますので正確に記述してください．空白は全て半角スペースです．なお，このカラムに限り省略可能です．

登録済み単語を検索する簡易検索ボックス

単語または読みの一部もしくは全部

入力できる文字の制限はありませんが，単語や読みとして登録できません．文字で検索をしても検索結果は得られませんので注意してください．(たとえば半角英字で検索した場合，検索対象には半角英字は絶対に含まれないため，どのような登録があった場合でも一件もヒットしません)

登録できる品詞

登録する品詞は, 正確に指定されるのが望ましいですが, 登録数が多い、または多分類が不明であるなどの場合は, 全て「名詞 一般」で登録しても検索には支障はありません. 但し, 正確な品詞で登録すると若干検索の精度が向上します. また, 品詞に含まれる空白文字は「半角の空白文字」1 つです. また, 前後に空白やタブ文字等が含まれていてはいけません. 後ほど解説する一括登録用の CSV を作成する場合はご注意ください.

名詞 一般

普通名詞, あるいは, 下位分類が未定の名詞

名詞 固有名詞

姓と名に分けられないもの, あるいは, 姓･名の決定が未定の人名

例:「お市の方」

名詞 固有名詞 人名 姓

主に, 日本人の姓.

例:「山田」

名詞 固有名詞 人名 名

主に, 日本人の名

例:「太郎」

名詞 固有名詞 組織

組織を表す名称

例:「通産省」「NHK」

名詞 固有名詞 地域 一般

国名以外の地名のなどを表すもの

例:「アジア」「バルセロナ」「京都」

名詞 固有名詞 地域 国

国の名前

例:「日本」「オーストラリア」

名詞 代名詞

いわゆる代名詞, 不定語

例:「あれ」「それ」「ここ」「どこか」「みなさん」「わたくし」

名詞 サ変接続

格要素を取り,「する」「できる」「なさる」「くださる」などが後接することのできるもの

例:「インプット」「愛着」「悪化」「悪戦苦闘」「下取り」

名詞 形容動詞互換

いわゆる形容動詞語幹で,「な」の前に現れるもの

例:「健康」「安易」「駄目」

2章
辞書の編集 その1
単語の登録, 削除

単語の個別登録

単語追加登録 メニュー

ここでは，弊社の社名である「タイムインターメディア」を「社名」として単語を一つ登録する場合を例にとって解説します．この「タイムインターメディア」という語は，通常の状態ですと「タイム」「インター」「メディア」と3つの別々の単語が並んでいるだけと扱われてしまうため，「タイム誌」「インターネット」「ウィンドウズメディアプレイヤー」などのように関連する並びの含まれた文書が検索結果に含まれてしまい，検索精度が落ちてしまいます．「タイムインターメディア」をひとつの単語として認識させ，検索精度が向上するよう単語登録してみます．

新規に単語を登録するには，「単語追加登録」のメニューを選択します．

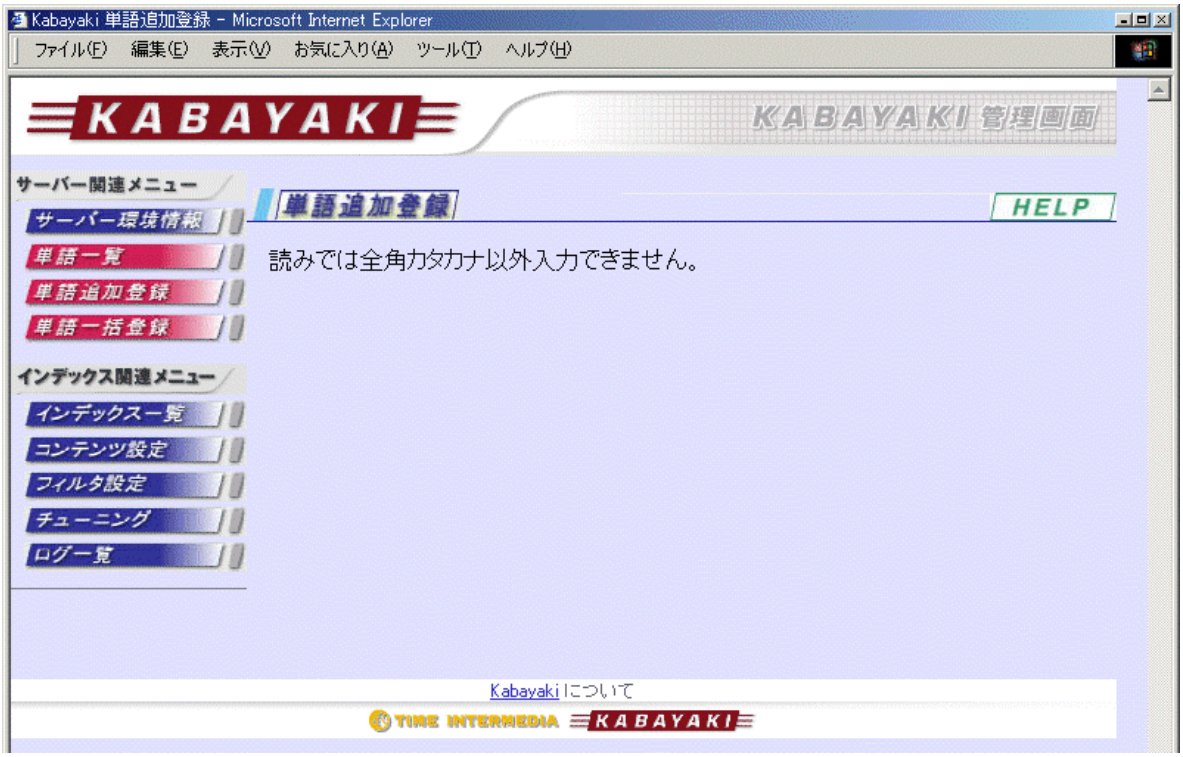
「単語追加登録」のメニューを選択すると，以下の画面が表示されます．

この画面では，以下の 3 つの入力･選択項目が存在します．

- 単語
- 読み
- 品詞

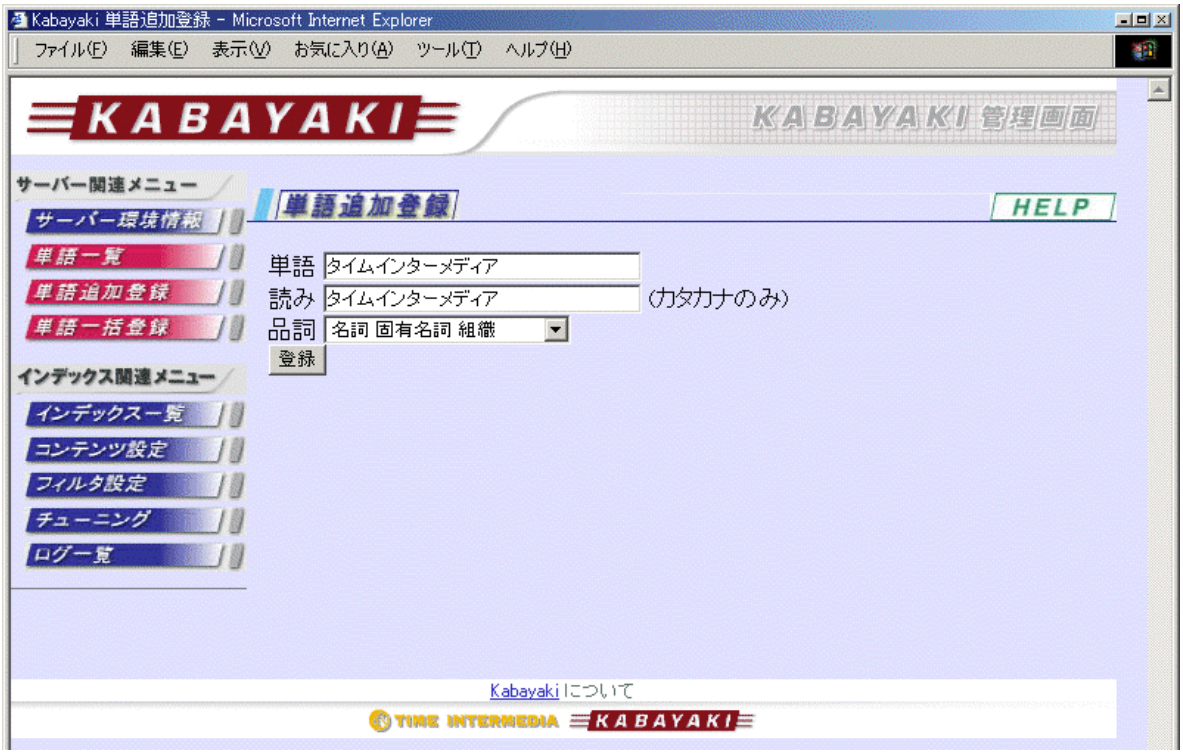
「単語」には，Kabayaki に一語として認識させたい語を入力します．今回は「タイムインターメディア」という単語を登録するので，「タイムインターメディア」と入力します．
ここで入力できるのは，漢字･平仮名･片仮名･全角英数字のみです．これら以外を入力するとエラー画面が表示されますので，画面の指示に従って再度入力してください．

「読み」には，単語の読みを片仮名で入力します．今回は「タイムインターメディア」という単語を登録するので，読みを「タイムインターメディア」とそのまま入力します．たとえば「事業推進本部」を登録したい場合は「ジギョウスイシンホンブ」となります．
ここで入力できるのは片仮名のみです．片仮名以外が含まれている場合はエラー画面が表示されますので，画面の指示に従って再度入力を行ってください．



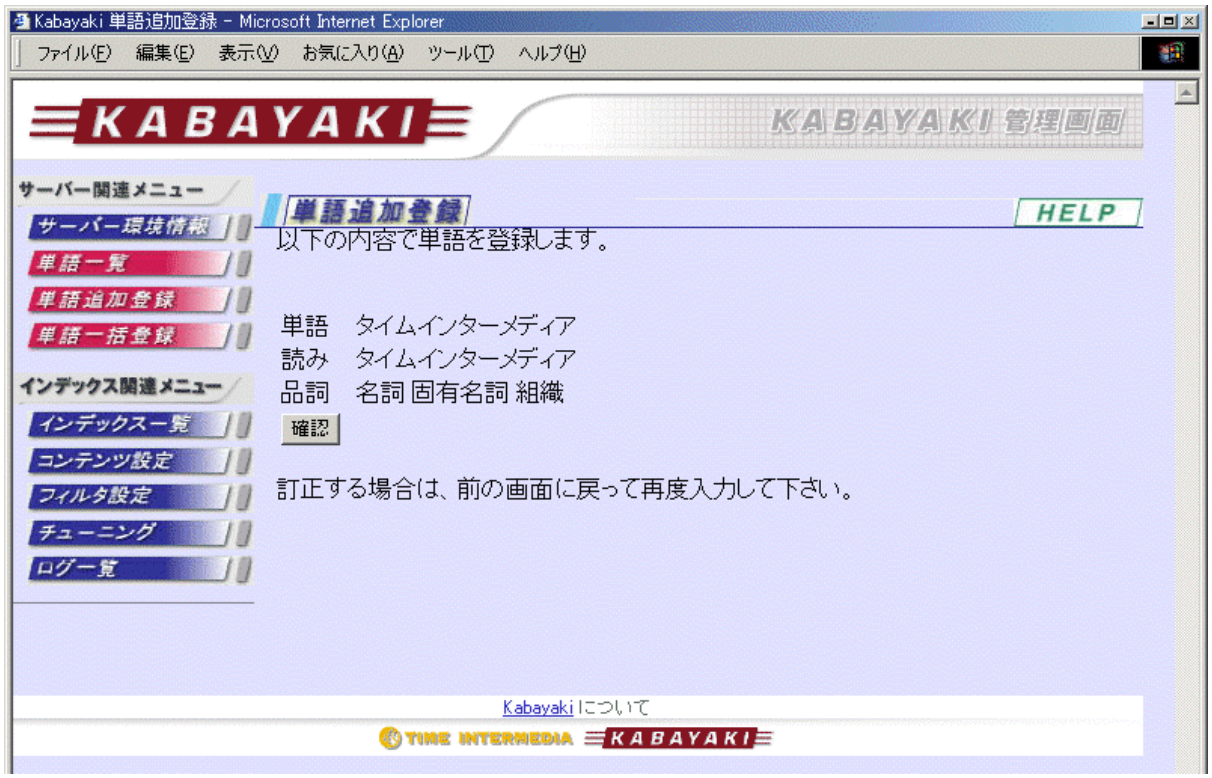
「品詞」では，登録する単語の品詞をプルダウンメニューにより選択します．今回登録する「タイムインターメディア」は社名ですので「名詞 固有名詞 組織」を選択します．ここを正確に選択することにより，若干，検索の精度が向上します．何を選んだらよいか解らない場合は，「名詞 一般」を選択しておけばよいでしょう．

選択できる品詞についての詳しい説明は，『登録できる品詞』を参照してください．



全ての項目を入力･選択したら,「登録」のボタンをクリックします.

入力内容の確認を行う画面が表示されます．誤りがなければ「確認」ボタンをクリックします．



もし確認画面ではなくエラー画面が表示される場合は，画面の指示に従って入力してください．入力可能な文字に関する詳しい説明は，『使用可能な文字』を参照してください．

登録内容の確認を実行すると，登録完了画面が出力します．

この登録は，次回のインデクシング(インデクシングに関しては Kabayaki 本体マニュアル参照)以降に有効になります．

単語の一括登録

単語一括登録メニュー

一度にたくさんの単語を登録したい場合，『単語追加登録』インターフェイスでひとつずつ行うのは手間がかかるでしょう．

Kabayaki では，まず，Microsoft Excel で単語の一覧を作成し，それを単語一括登録メニューからアップロードすることにより，容易に複数の単語を一括で登録することが可能です．

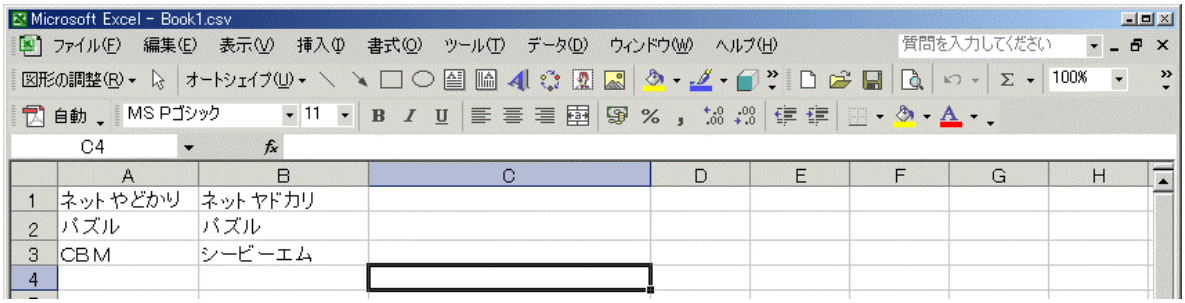
まず，Microsoft Excel を用いて，以下の書式で一括で登録する単語の一覧を[1]作成します．

単語,読み,品詞

なお，Microsoft Excel の操作に関する詳細は，付属の取扱説明書をご覧ください．

単語名や読みに使える文字の制約に関しては，『使用可能な文字』を参照してください．

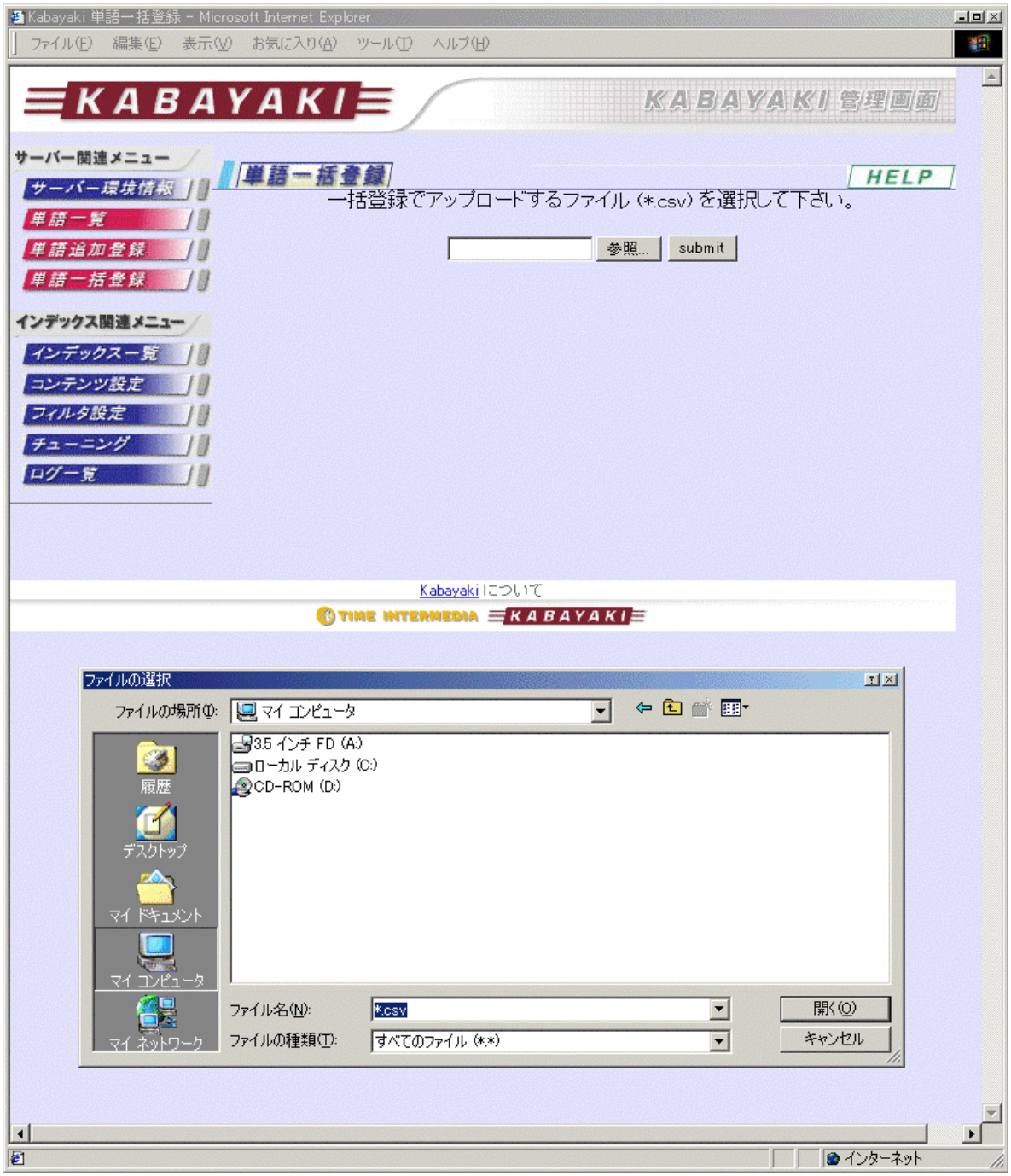
品詞は省略可能です．品詞を省略した場合は「名詞 一般」となります．品詞に関しては，一文字漏らさず正確に記入しないとエラーになります．まずは省略して登録し，後ほど解説します，「更新」の機能で「名詞 一般」ではないものだけを選び出して，正しい品詞を割り当てるとよいでしょう．



[1] Microsoft Excel でファイルを保存する際には，『CSV形式』で保存してください．CSV形式以外で保存した場合，Kabayaki辞書管理ツールで単語一括登録することができません．

次に Kabayaki 辞書管理ツールより,「単語一括登録」を選択します.

次に，さきほど作成した CSV ファイルを，ウェブブラウザ(Internet Explorer, Netscape など) を用いてアップロードします．下図で，参照... と submit とあるボタンは，お使いの WWW ブラウザによって表記が異なりますので注意してください．



これで，単語の一括登録の処理が完了し，実行結果が表示されます．画面には，単語の読み取り件数，登録の成功件数と失敗件数が表示されます．登録を失敗したものに関しては，どの単語が失敗したかが一覧されます．一括登録に失敗した単語に関しては，CSV ファイルを修正の上再度登録してください．

全ての単語が登録できた場合

登録がエラーになった場合

単語の削除

単語のリスト表示画面より，登録済みの単語の削除をおこないます．
たとえば，誤って「本日は晴天なり」という単語を追加してしまい，削除したい場合は，以下のようにします．

まず，単語のリスト表示で「本日は晴天なり」を表示させます．表示させる方法は，後ほど解説します『単語一覧』の検索の機能を使うとよいでしょう．

次に，一覧の左端の数字をクリックします．

次に，削除ボタンをクリックします．

続いて，削除の確認を求める画面が表示されますので，削除してよいか確認してください．よろしければ「確認」ボタンをクリックします．

これで，作業完了画面が出力され，単語の削除が完了します．

何らかの理由で単語の削除に失敗した場合は, その旨画面に表示されますので, 指示に従って引き続き操作してください.

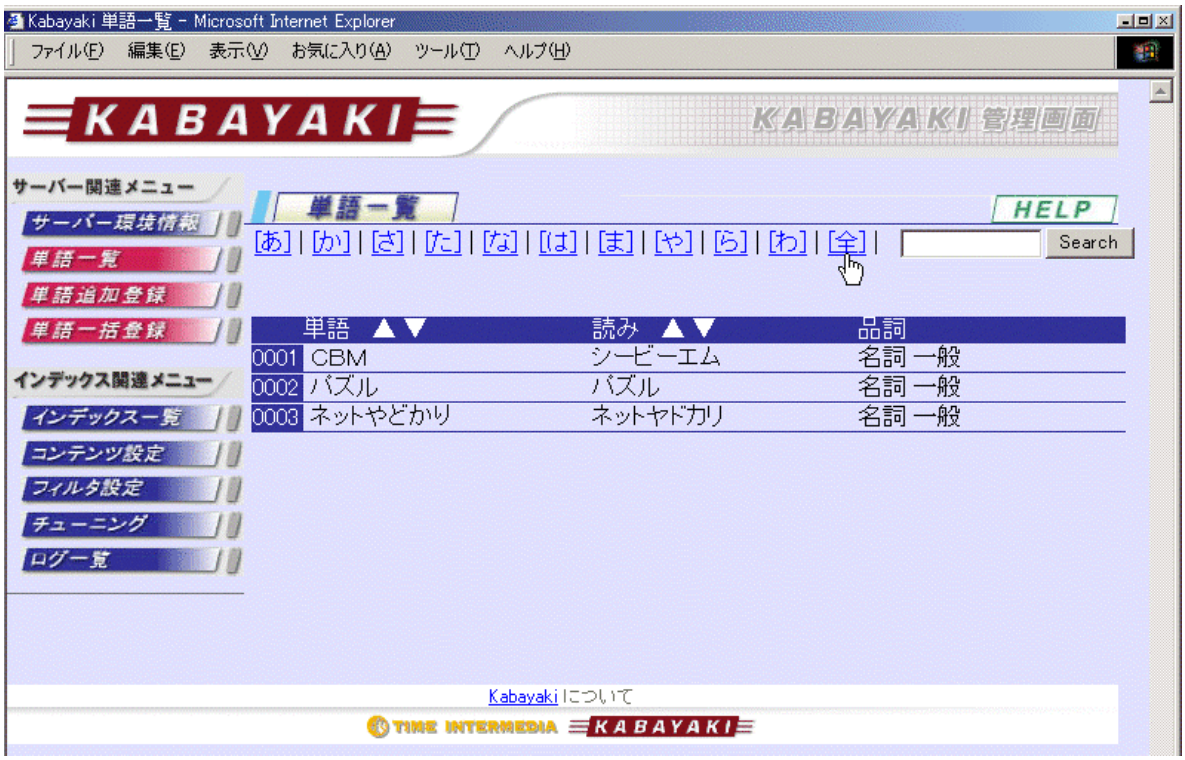
3章
辞書の操作
単語の検索，更新

単語の検索･一覧表示

辞書を検索することで，既に登録されている単語を確認し，新しい単語の登録と必要のない単語の削除を手助けします．

登録されているすべての単語を一覧する場合

「検索･単語リスト」の画面から「全て」を選択します．



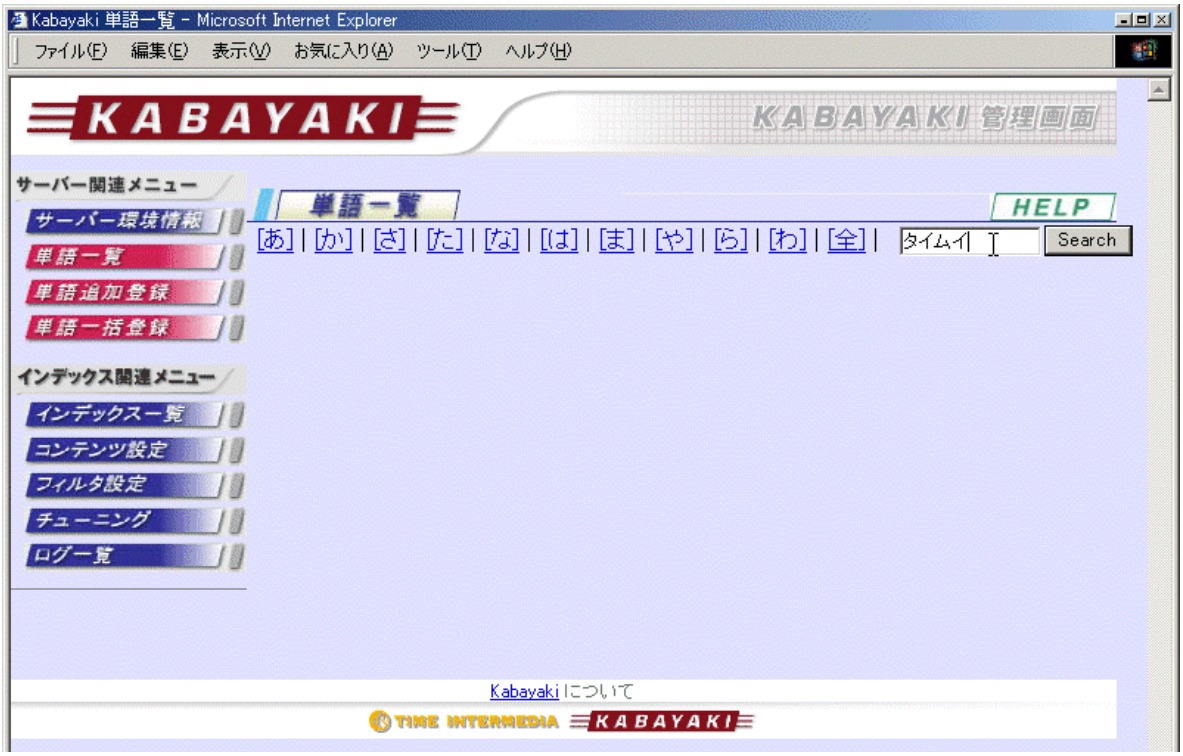
登録した単語の頭文字で検索し一覧する場合

「検索･単語ﾘｽﾄ」の画面から，単語の読みの頭文字を選択します．たとえば，タイムインターメディアを

確認する場合は「た」を選択します．

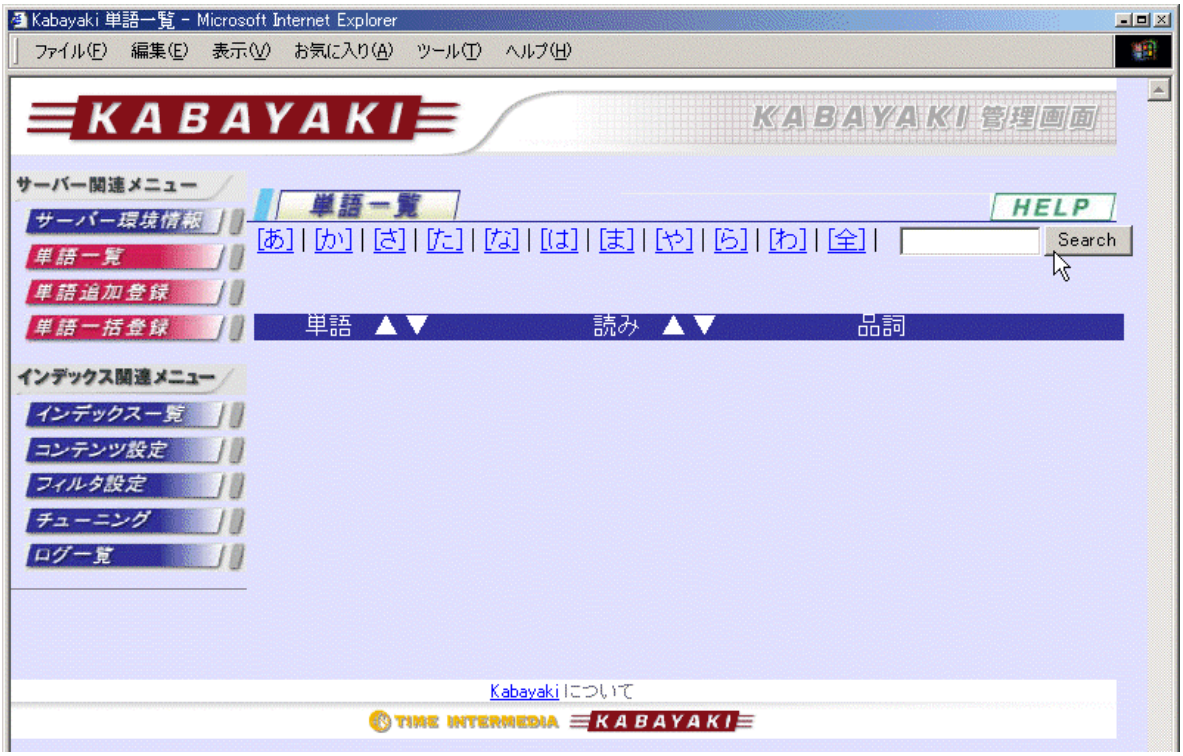
単語の読みで検索する場合

右上段検索ボックスに単語の読みの一部，もしくは全部を入力します．



いずれの方法で検索した場合も，以下の形式で単語の一覧が表示されます．条件に該当する単語が存在しない場合は，その旨のメッセージが画面に表示されます．

条件に該当するものが存在した場合



条件に該当するものが存在しなかった場合

単語の更新

単語の一覧表示画面より，登録内容の更新ができます．

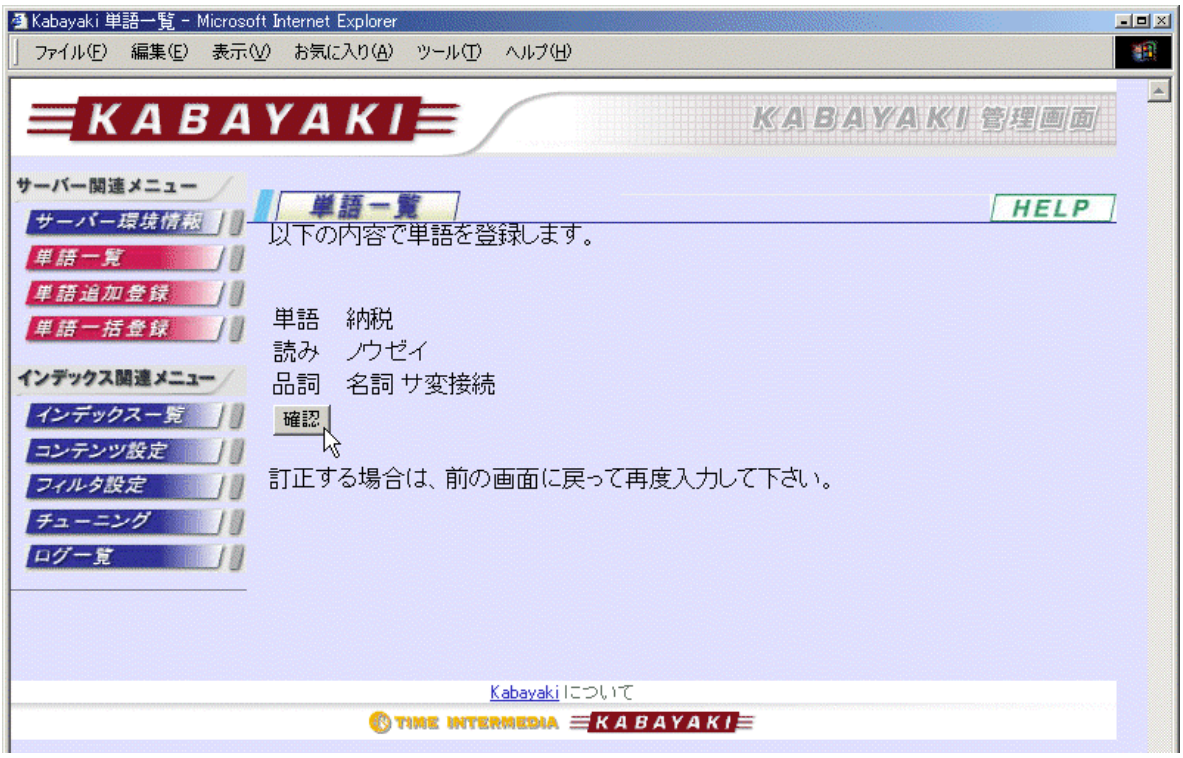
ここでは，「納税」という単語を「名詞　一般」で登録していて，これを「名詞　サ変接続」に修正する場合は例にとって解説します．

まず，単語一覧で「納税」を表示させます．単語の表示についての詳しい説明は，『**単語の検索･一覧表示**』を参照してください．

次に，一覧の左端の数字をクリックします．

次に，変更したい部分を入力し，「登録」ボタンをクリックします．

次に，確認画面にて，更新の内容が正しいことを確認して，「確認」ボタンをクリックします．



これで，作業完了画面が出力され，単語の更新が完了します．

何らかの理由で単語の更新に失敗した場合は，その旨画面に表示されますので，指示に従って引き続き操作してください．